MOVIE RECOMMENDER SYSTEM

Presented by joyce muthiani

Table of contents

Overview

The objective of this project is to develop a movie recommendation system that tailors suggestions to individual users, taking into account their past viewing habits and preferences. The system utilizes a combination of collaborative and content-based filtering methods to improve the precision and pertinence of the movie recommendations.

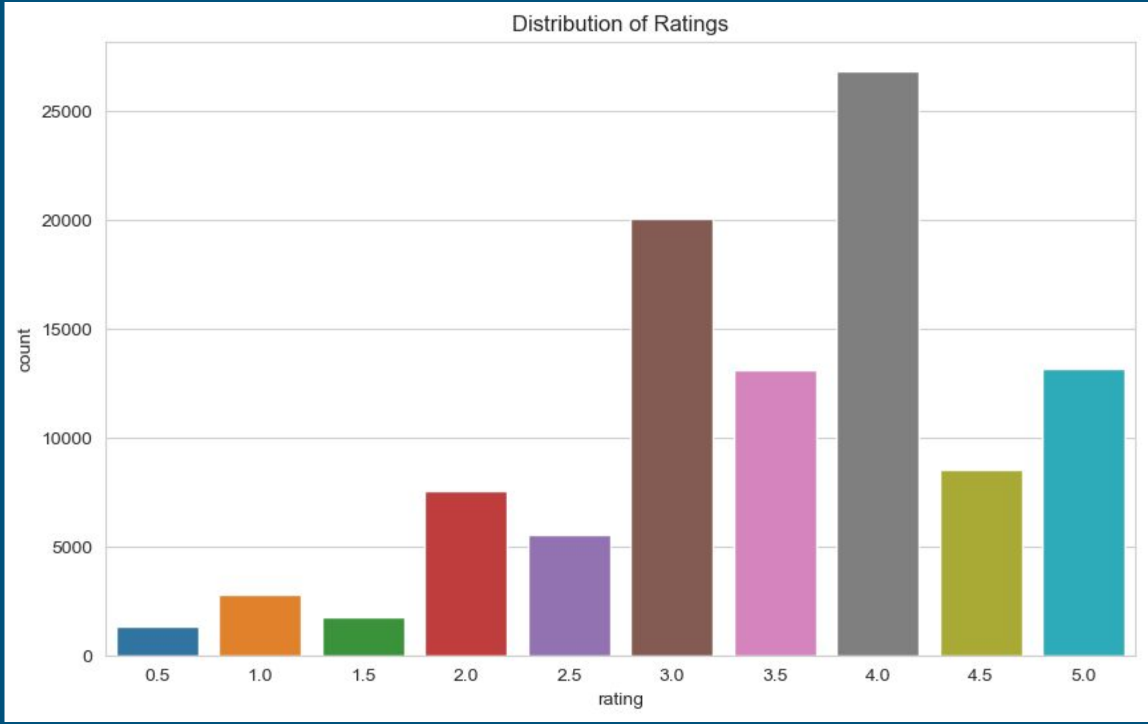
Business Understanding

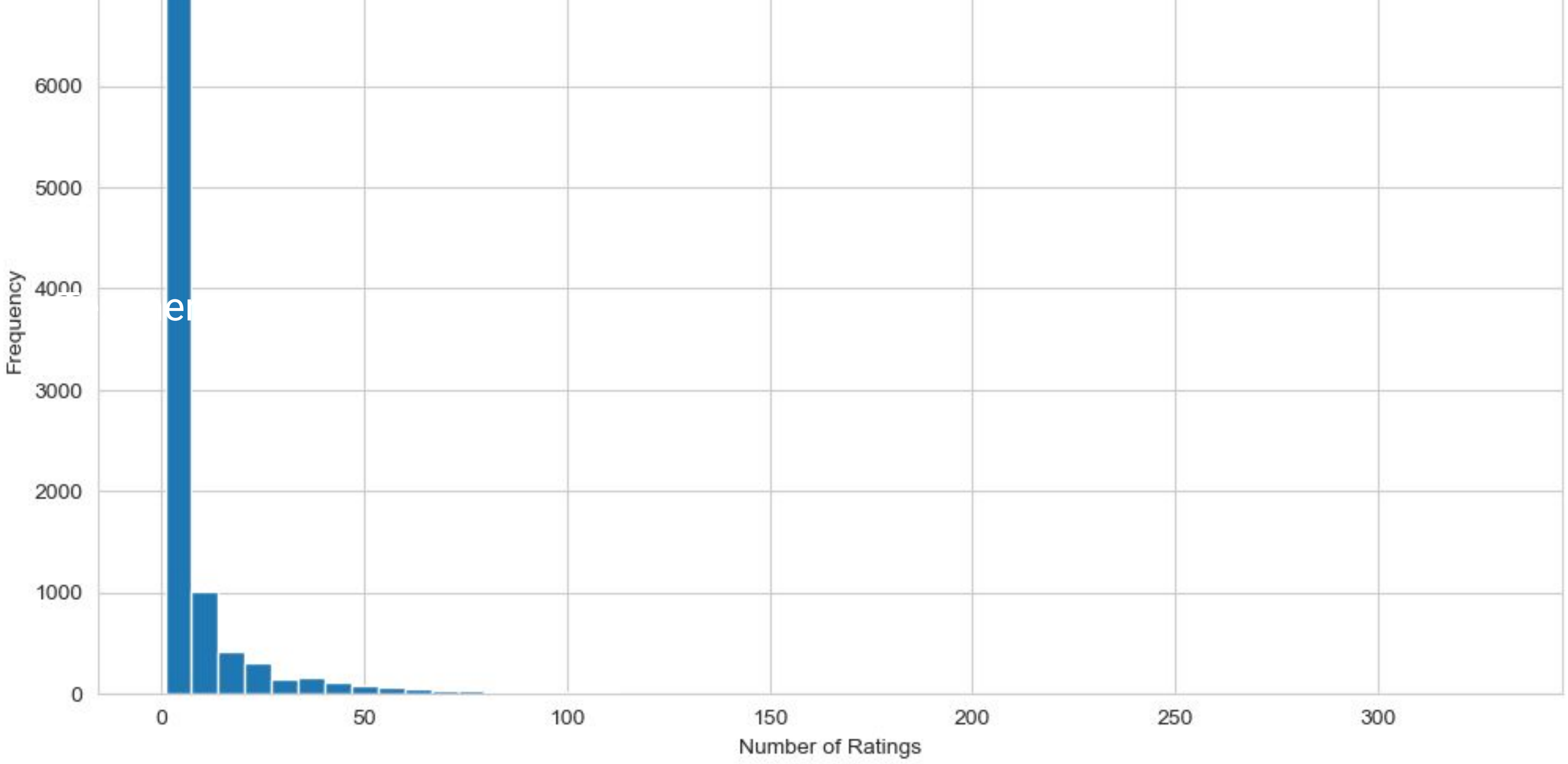
The main goal of this project is to boost user engagement and satisfaction on the MovieLens platform by providing personalized and relevant movie recommendations.

The recommender system is designed to offer customized suggestions to users based on their historical movie ratings and tagging activities, thereby enhancing their overall user experience.

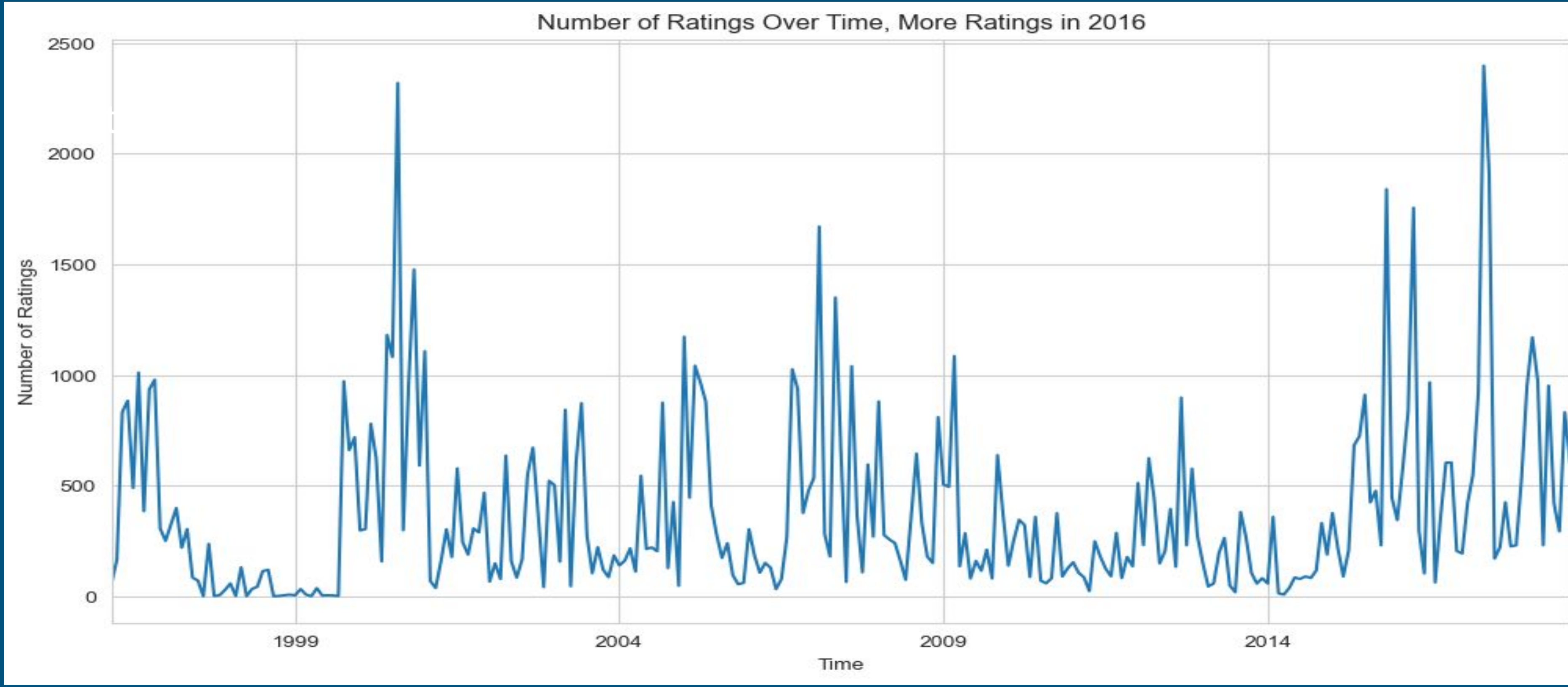
Visualizations

A c o u n t plo t vis u alizin g t h e dist rib u tio n o f r a tin gs a c r oss m o vie se




6000
5000
4000
3000
2000
1000
0
Frequency
0
50
100
150
200
250
300
Number of Ratings

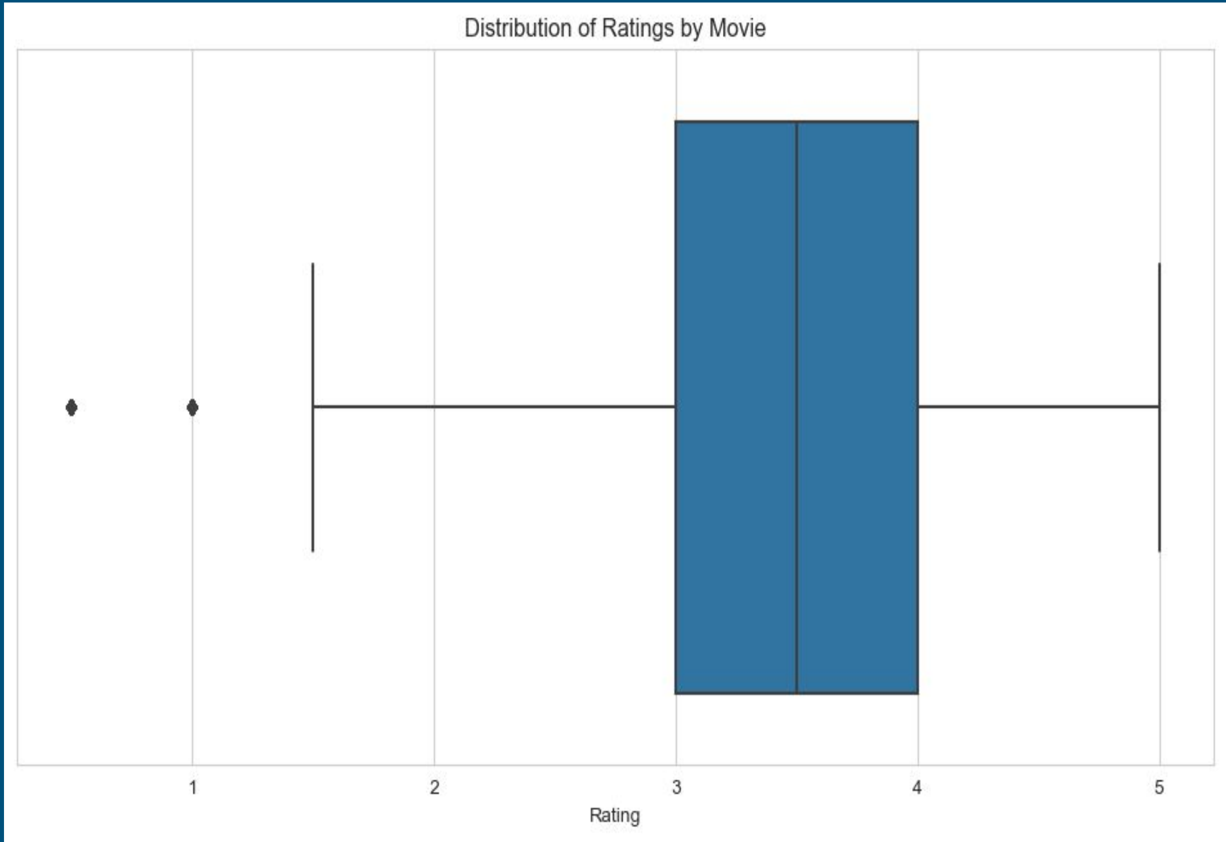
Number of Ratings Over Time,More Ratings in 2015



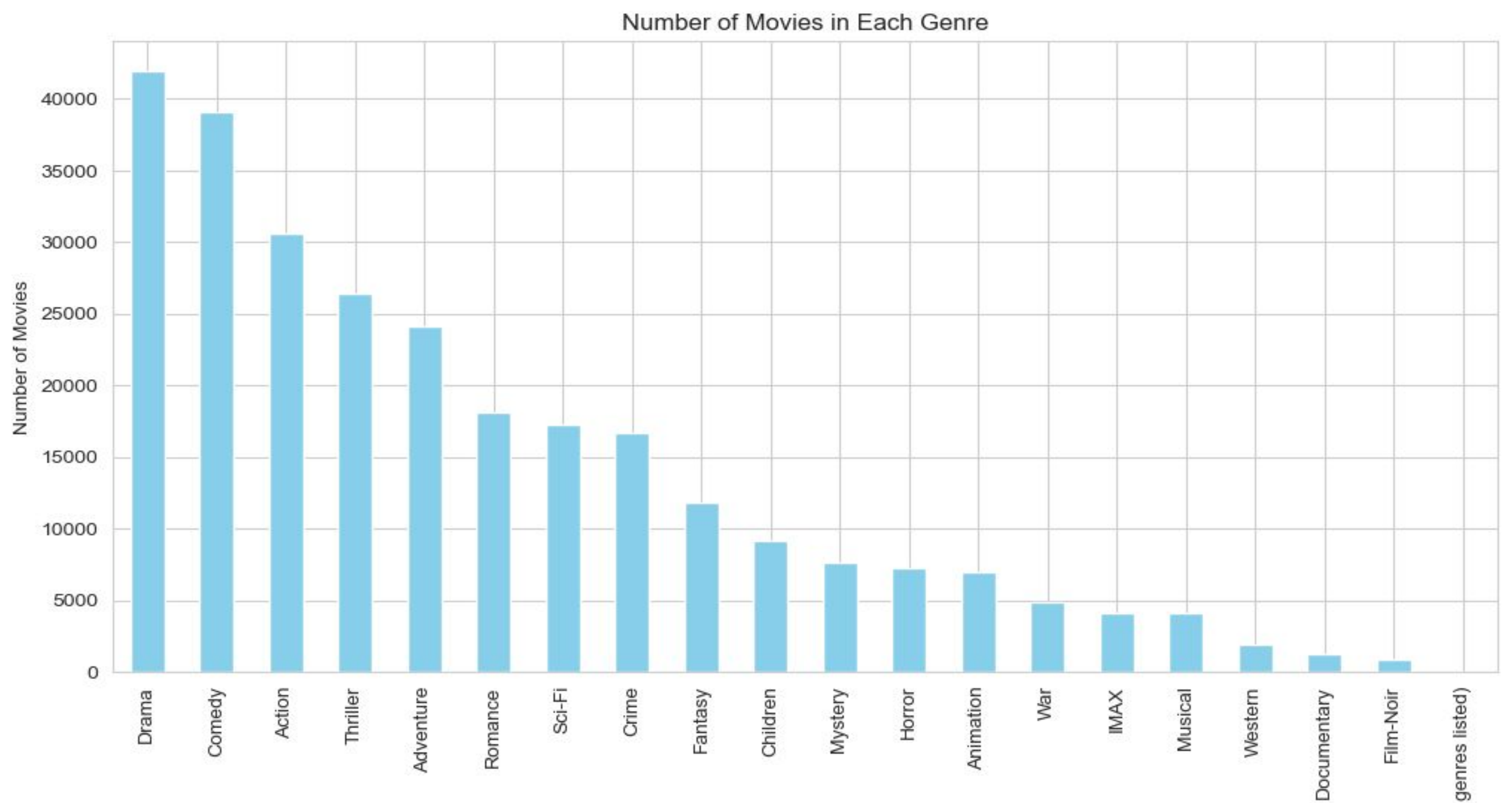
A wordcloud representing movie genres

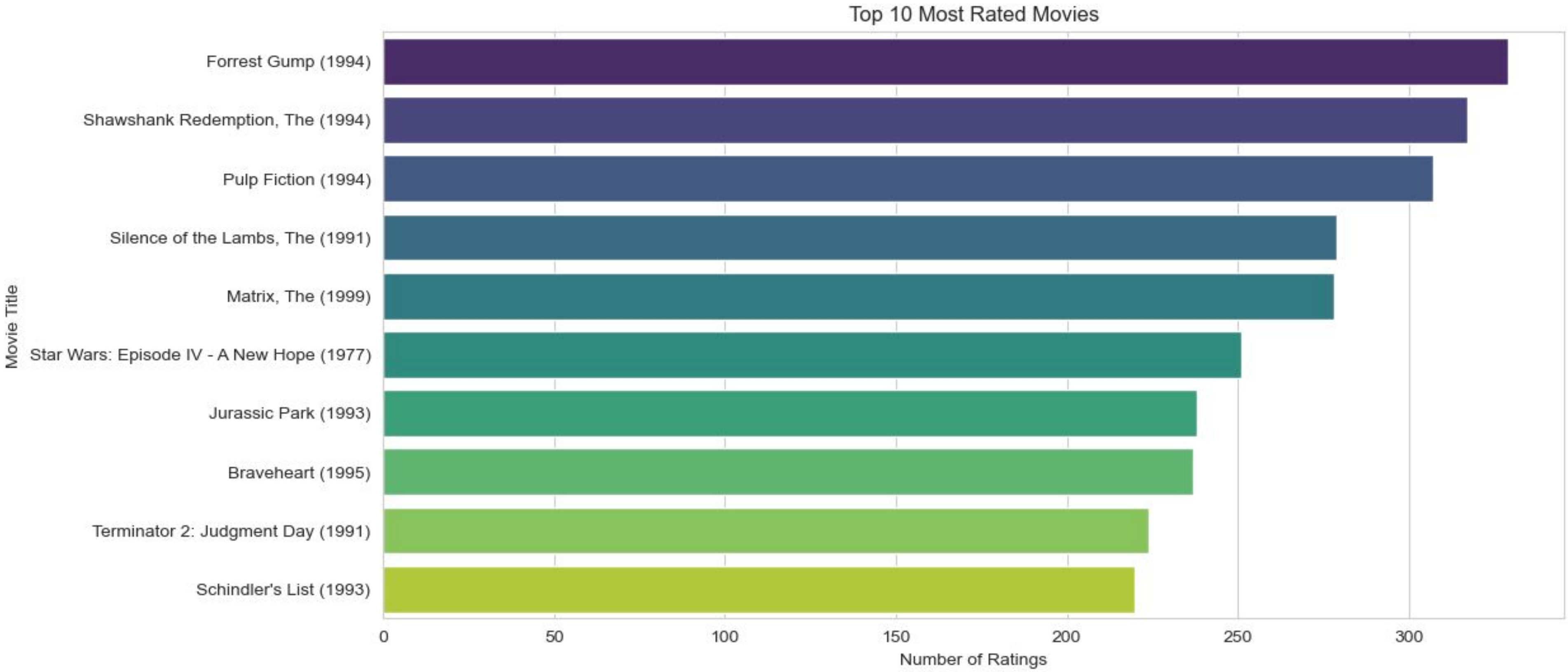
A box plot Visualizing the Distribution of Ratings by Movie

Testx here



A bar chart visualizing the distribution of movies across different genres





Top 10 Most Rated Movies
Forrest Gump (1994)
Shawshank Redemption, The (1994)
Pulp Fiction (1994)
Silence of the Lambs, The (1991)
Matrix, The (1999)
Star Wars: Episode IV - A New Hope (1977)
Jurassic Park (1993)
Braveheart (1995)
Terminator 2: Judgment Day (1991)
Schindler's List (1993)
Movie Title
0
50
100
150
200
250
300
Number of Ratings

Modelling

1. Content - Based Filtering
2. Neighbourhood - Based collaborative filtering
3. Model - Based Methods - Matrix Factorization using SVD

Summary of the models

1.Content - Based Filtering

Strengths: Relies on item features for recommendations.

Limitations: May struggle with diverse user preferences.

2.Neighbourhood - Based collaborative filtering

Advantage: The neighborhood-based collaborative filtering model outperforms the content-based approach by leveraging user ratings for a deeper dataset analysis and capturing nuanced patterns to enhance recommendation accuracy.

Disadvantage: The cold start problem poses a concern for new users or items with insufficient data, while scalability issues may emerge with dataset growth, necessitating efficient algorithms and ample computing resources for sustained performance.

3. Model based methods

Strengths: Utilizes matrix factorization for personalized recommendations.

Limitations: Requires hyperparameter tuning and may face the cold start problem.

Coclusion

Content-Based Filtering: I constructed a content-based recommender by leveraging movie attributes like genres. This approach excelled in suggesting similar movies based on content similarities. However, it may struggle with capturing diverse user preferences and recommending outside the established content boundaries.

Neighborhood-Based Collaborative Filtering (KNN): Implementing KNN models using SciKit Learn, I tapped into user-item interactions to drive recommendations. The item-based variant, utilizing cosine similarity, showcased notable effectiveness in identifying analogous movies, enriching the user experience.

Model-Based Collaborative Filtering (SVD): my experimentation extended to the Surprise library, where I employed SVD, a matrix factorization method. While SVD demonstrated reasonable performance with an RMSE of 0.8925, its efficacy hinged on fine-tuned hyperparameters and meticulous model assessment.

RECOMMENDATIONS

Considering the nuanced trade-offs inherent in each method, I advocate for a hybrid model amalgamating content-based and collaborative filtering. This hybrid framework can harness the granular user-item interactions from collaborative filtering while integrating content cues for enhanced personalization and diversity in recommendations.

Furthermore, I propose ongoing hyperparameter optimization and model refinement, particularly with larger datasets, to augment the predictive power of collaborative filtering techniques. Continual feedback integration from users and dynamic content updates will further fortify the recommendation engine's efficacy and relevance over time.

In essence, the optimal recommender system choice hinges on specific use-case nuances, user dynamics, and dataset intricacies. A meticulously curated hybrid model, iteratively honed and validated, stands poised to deliver robust, accurate, and adaptive movie recommendations, aligning seamlessly with evolving user preferences and content landscape dynamics.